B. Specification

Please amend paragraph [0004] as follows:

--With the conventional selection method for selecting a portion with a high level of uniqueness described above, uniqueness evaluation is performed with regard to general sequences. For example, in the event of creating a DNA microarray for DNA from a human genome, uniqueness was checked for all human genome base sequences, and a portion with a high level of uniqueness was selected as a probe base sequence.--

Please amend paragraph [0005] as follows:

--However, there has been a problem with the conventional selecting method in that, in the event that extremely similar base sequences are contained in a specimen, and the similar base sequences include base sequences belonging to one group and base sequences belonging to another group, determining whether each base sequence belongs to that group is extremely difficult. More specifically, at the time of making a determination for an infection or the like, there has been the problem that it is extremely difficult to find a probe which exhibits the same degree of hybrid strength regarding a DNA base sequence of the same strain of bacteria and which exhibits a different degree of hybrid strength regarding a DNA base sequence of another strain of bacteria.--

Please amend paragraph [0054] as follows:

--Normally, PCR amplification is performed so as to amplify the portion of a base sequence making up a ribosome RNA called 16s (16s rRNA) in the event that

determining an infectious bacteria is the object. In this case, the PCR primer for the *staphylococcus aureus* shown to the left in Fig. 4, and the PCR primer for the *escherichia coli* to the right side, are almost the same. More specifically, multiplex PCR using a primer set capable of amplifying the locus coding 16s rRNA of any bacterium is preferable. In this case, both the left and right hybridization solutions (404, 414) in Fig. 4 consequently contain multiple types of base sequences. The reason for this will be described in detail with reference [[with]] to the subsequent drawing.--

Please amend paragraph [0056] as follows:

--Now, in the event that the DNA microarray designed for determining the *staphylococcus aureus* works correctly, the spot will ~~reactive~~ react positively in the hybridization solution 404 (405)[[,]] and react negatively in the hybridization solution 414 to the right (415). In the same way, in the event that the DNA microarray designed for determining the *escherichia coli* works correctly, the spot will ~~reactive~~ react negatively in the hybridization solution 404, and react positively in the hybridization solution 414. Of course, the bacterium may be determined using a DNA microarray wherein several types of spots, each reacting uniquely to different bacteria, are arrayed.--

Please amend paragraph [0057] as follows:

--Next, the reason why multiple types of base sequences exist in the hybridization solution in Fig. 4 will be described with reference to Fig. 5. Bacteria in the natural world tend to frequently mutate. As a result, there may be multiple types of major

strains, which have survived natural selection, coexisting simultaneously. For example, bacteria strains [[which]] that cause a so-called "hospital infection" emerge [[by]] from a bacterium, which normally has no drug resistance, mutating and consequently acquiring drug resistance. Once such drug resistance is acquired, the bacterium may manifest itself having robust ~~reproductively~~ reproductive ability even in sanitary environments, which are aggressively sterilized. Thus, it is proper to assume that there are several variations to each base sequence of any bacterium [[which]] that exists in the natural world.--

Please amend paragraph [0059] as follows:

--In order to achieve this object, with the probe design method according to the present embodiment, frequency tables are compiled separately for a set of base sequences that belong to the same group as the target base sequence (own base sequence data 101) and a set of base sequences that belong to a group competing with the former group (competing base sequence data 103) as shown in Fig. 1. In the example shown in Fig. 4, a collection of base sequences of 16s rRNA at various loci from various strains of staphylococcus aureus make up the own base sequence data 101, and a collection of base sequences of 16s rRNA at various loci from various strains of bacteria other than staphylococcus aureus, such as *escherichia coli* and *haemophilus influenzae*, make up the competing base sequence data 103.--

Please amend paragraph [0060] as follows:

--The way in which a frequency table is compiled from such a base sequence is shown in Fig. 6. To compile a frequency table, the number of times ~~of occurrence of~~ a partial sequence of a length "n" (in Fig. 6, n=9) is present in the base sequence data is counted. The variations of [[a]] an n-long base sequence ~~n long~~ is 4 to the n'th power, so in Fig. 6, the number of lines is 4^n. Note that in Fig. 6, the lower the frequency of emergence [[is]], the higher the uniqueness of the partial sequence [[is]], so the frequency multiplied by minus 1, for example, represents the uniqueness.--

Please amend paragraph [0061] as follows:

--That is to say, in the own frequency table creating step 102, reference is made to the own base sequence data 101 storing the base sequences for 16s rRNA of various strains of the bacterium to be detected, the number of ~~times of occurrence~~ occurrences is counted for all partial base sequences having a length of n, and the results are compiled in a table as shown in Fig. 6, thereby creating the own frequency table 105. In the same way, in the competing frequency table creating step 104, reference is made to the competing base sequence data 103 storing the base sequences for 16s rRNA of various strains of ~~bacterial~~ bacteria other than the bacterium to be detected (i.e., a bacterium to be distinguished from the bacterium to be detected), the number of ~~times of occurrence~~ occurrences is counted for all partial base sequences having a length of n, and the results

are compiled in a table as shown in Fig. 6, thereby creating the competing frequency table 106.--

Please amend paragraph [0064] as follows:

--For example, in the case of the graph shown in Fig. 8, a portion with a high level of uniqueness of the target base sequence exists at a portion around 2/3 from the head of the sequence, i.e., at a position around 1000 bases down. For example, in the event that this target base sequence is the first 16s portion (array X) of the Mu50 strain shown in Fig. 5, this means that a base sequence unique to this array exists around 1000 bases down of the sequence X, which is a sequence different from other 16s portions on the Mu50 strain and different from the 16s rRNA from strains of *staphylococcus aureus* other than Mu50. Accordingly, selecting a probe candidate from this portion is unsuitable. Also, the portions with a low uniqueness in the upper graph ~~means~~ indicate that the bacterium cannot be distinguished from other bacteria, so selecting a probe candidate from such portions is unsuitable.--

Please amend paragraph [0066] as follows:

--Consequently, a probe is selected which exhibits a strong hybridization reaction for the same bacterium regardless of the loci and strains of the 16s rRNA coding sequence contained in the hybridization solution, and which exhibits a weak hybridization reaction for different bacteria regardless of the loci and strains of the 16s rRNA coding sequence contained in the hybridization solution.--

Please amend paragraph [0067] as follows:

--Note that the probe design method according to the present invention is not restricted to ~~application to determining~~ applications aimed at identifying infections; ~~rather~~. Rather, the method can be applied to any case wherein there is some degree of variation in a base sequence generally judged to be the same. For example, this may be applied to MHC widely used for individual identification of humans, and so forth.--

Please amend paragraph [0089] as follows:

--With the first embodiment, the frequency information is displayed as shown in Fig. 11, so as to allow the user to select suitable positions. Using the frequency information in this way enables the user to easily select suitable probe candidates, but the number of probes set for microarrays is generally [[great]] large, [[in]] on the order of hundreds if not thousands. Accordingly, an arrangement wherein a user sets all of the probes based on frequency information can require a great amount of time and trouble. Also, as stated in the first embodiment, partial base sequences can be automatically extracted by simply comparing uniqueness values with threshold values. However, in this case, there are problems that (1) searching over the entire base sequence length requires a long time for calculations, (2) there is the possibility that a great number of similar base sequences may be extracted, and (3) there is a difficulty in extracting partial base sequences from positions suitably dispersed over the entire length of the base sequence.--

Please amend paragraph [0091] as follows:

--Fig. 12 is a flowchart description of the procedures for the probe design method according to the second embodiment. The steps and data [[which]] that are the same as those in the first embodiment (Fig. 1) are denoted with the same reference numerals.--

Please amend paragraph [0092] as follows:

--Reference numeral 1201 denotes all base sequence data, which is a collection of the own base sequence data 101 and the competing base sequence data 103. Reference numeral 1202 denotes a common sequence data creating step for extracting partial sequences common ~~with the~~ to all base sequence data, and creating common sequence data 1203. ~~note that the~~ The common partial base sequences are base sequences of a predetermined number of bases or longer (e.g., base sequences with a length of 20 bases or more), and are obtained by searching all base sequences.--

Please amend paragraph [0094] as follows:

--Now, automatic probe design according to the present embodiment will be described. With the present embodiment, the common region data 1203 is used for the automation of the probe design. The common region data 1203 is created in the common region data creating step 1202, where [[the]] all base sequence data 1201, which is a collection of the own base sequence data 101 and the competing base sequence data 103, is searched for partial sequences common to all base sequences, and the position on the

sequence and the length thereof are saved as common sequence data 1203. With the example of bacterial 16s rRNA, the common partial sequences are known to be at similar positions.--

Please amend paragraph [0095] as follows:

--Making reference the common sequence data 1203 with regard to the target base sequence data 107 allows the common regions 1302, 1303, and so on through 1306 and so forth, and the regions 1303, 1313 and so on through 1315 and so forth, between the common regions, to be distinguished on the target base sequence denoted by reference numeral 1301, as shown in Fig. 13. One position wherein the uniqueness between strains of the same bacterium is low and the uniqueness between different bacteria is high[[,]] is selected by making reference to the own frequency table 105 and the competing frequency table 106 in the regions between the common regions. There are multiple common regions 1302 on the target base sequence 1301, and accordingly multiple regions 1303 between the common regions, so probes can be set over the entire length of the target base sequence 1301 by mechanically repeating the same process as long as there are unprocessed regions between the common regions. This processing can also be mechanically processed even in the event that there are a great number of target base sequences 110 over a range of multiple types of bacteria, and accordingly can be automated by a computer.--

Please amend paragraph [0096] as follows:

--Note that, as [[with]] in the first embodiment, the probe design method according to the present invention is not restricted to ~~application to determining~~ applications aimed at identifying infections~~, rather~~. Rather, the method can be applied to any case wherein there is some degree of variation in a base sequence generally judged to be the same. For example, this may be applied to MHC widely used for individual identification of humans, and so forth.--

Please amend paragraph [0098] as follows:

--The flow of the probe design program according to the second embodiment will be described with reference to Figs. 14 through 16. In Fig. 14, the processing and data that are the same as in the first embodiment (Fig. 9) are denoted with the same reference numerals. As described above with reference to Fig. 9, in the target organism group selection (901), genome information regarding, for example, ~~such as~~ bacterium, virus, fungus, and the like, belonging to a target organism group selected according to the probe to be designed, is selected from a base sequence database 906, and stored in the garget base sequence database 907.--

Please amend paragraph [0101] as follows:

--Also, common region data 913 (equivalent to the common region data 1203 in Fig. 12) is created (921) along with the frequency tables (903). Information of the partial sequences (common region data 912) shared between all base sequences (912)

contained in the target base sequence database 907 is stored in a common region table 913. Compiling of the frequency tables is the same as described [[with]] in the first embodiment.--

Please amend paragraph [0106] as follows:

--Fig. 16 illiterates an example of a design screen according to the preset embodiment. The interface configuration is approximately the same as in the first embodiment (Fig. 11), and the same components are denoted with the same reference numerals. Reference numerals 1101 through 1104 denote graphs, and the horizontal axis represents the position on the target base sequence selected with a user interface, such as shown, for example, in Fig. 15~~, for example~~. As described with reference to Fig. 11, the graphs 1102 and 1103 correspond to the upper and lower graphs in Fig. 8, with the graph 1102 showing the uniqueness of the partial sequence at each position as to the competing base sequence data, and the graph 1103 showing the uniqueness of the partial sequence at each position as to the own base sequence data. Also, graph 1101 shows the uniqueness of the partial sequence at each position as to the human genome. Graph 1104 shows the melting temperature of base sequences of a predetermined number of bases (in this example, base sequences of 24 bases) starting at each position.--

Please amend paragraph [0109] as follows:

--~~It is needless to say that a~~ A manual mode may be provided wherein a user manually sets probes (i.e., specifies partial base sequences. In this manual mode, probes should be specified at portions where the graph 1102 peaks and the graph 1103 shows a trough, as described [[with]] in the first embodiment (Fig. 8 or Fig. 11). In this case, an arrangement may be made wherein the head of the partial base sequences cannot be specified in common regions indicated by the points 1601.--

Please amend paragraph [0110] as follows:

--Reference numeral 1105 denotes an information space, for displaying the current target species, various types of parameters, and so forth. Note that 24 is set as the default base length with the present embodiment, and the melting temperature is calculated for the graph 1104 based on this. Reference numeral 1106 denotes a list of ~~deigned~~ designed probes, with the position thereof being displayed by a dotted line 1107. The solid line 1108 represents the "current" position, which is the position of interest as of now. The partial base sequence corresponding to that position (24-base base sequence) is displayed in the space 1109, the base sequence immediately prior to that position is displayed in the space 1110, and the base sequence immediately following that position is displayed in the space 1111. With the present embodiment, the sequences of the 10 bases before and after are displayed.--

Please amend paragraph [0114] as follows:

--~~Let as assume~~ In a case wherein there are five common regions, 1402 through 1406, on the target base sequence 1401. ~~First~~, first, a probe will be created for between the common regions 1402 and 1403, i.e., the region denoted by 1407. Regardless of whether ~~or not creating~~ the probe between the common regions 1402 and 1403 ~~succeeds~~ is successfully created, next, a probe is created for between the common regions 1403 and 1404 in the same way, followed by the common regions 1404 and 1405, and then the common regions 1405 and 1406, so probe fabrication is attempted in sequence at all regions between the common regions.--

Please amend paragraph [0139] as follows:

--The genome DNA of the microorganism (enterobacter cloacae strain) collected was subjected to agarose electrophoresis and 260/280 nm light absorption measurement, thereby inspecting the quality (amount of low-molecular nucleic acid contained and degree of decomposition) and amount collected according to the method.--

Please amend paragraph [0142] as follows:

--Next, the substrate was immersed for 10 minutes [[an]] in a 1N sodium hydroxide solution heated to 80°C. Purified water cleansing and ultrapure water cleansing were repeated, thereby preparing a quartz glass substrate to serve as a DNA chip.--

Please amend paragraph [0152] as follows:

--The DNA microarray fabricated in "4. Fabricating the DNA microarray" and the labeled specimen fabricated in "5. Amplification and labeling of specimen (PCR amplification and fluorescent label inclusion)" were used for the detection reaction.--

Please amend paragraph [0154] as follows:

--The spin-dried DNA microarray was set in a hybridization device (Hybridization Station manufactured by Genomic Solutions Inc.), and a hybridization reaction was carried out with the hybridization solution and under the conditions shown below.

- Hybridization Solution

  6×SSPE / 10% formamide / Target (all 2nd PCR products)

  (6×SSPE: 900 mM of NaCl, 60 mM of $NaH_2PO_4 \times H_2O$, 6 mM of EDTA, pH 7.4)

- Hybridization Conditions

  65° C 3 minutes → 92° C 2 minutes → 45° C 3 hours → Wash 2×SSC/0.1% SDS at 25 °C → Wash 2×SSC at 20 °C → (manually wash with H_2O) → spin dry--

Please amend paragraph [0155] as follows:

--That is to say, the hybridization reaction was carried out for 3 minutes at 65°C, 2 minutes at 92°C, and 3 hours at 45°C, and then cleansed with 2×SSC/0.1% SDS at 25°C and 2×SSC at 20°C, and finally rinsed with purified water and spin-dried.--

Please amend paragraph [0156] as follows:

--[[the]] The DNA microarray following the hybridization reaction was subjected to fluorescence measurement using a DNA microarray fluorescence detecting device (GenePix 4000B, manufactured by Axon Instruments, Inc.). Excellent discrimination results were obtained with each of the probes.--